



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genetic diversity and multiplicity of infection in *Fasciola gigantica* isolates of Pakistani livestock

Citation for published version:

Ur-Rehman, Z, Zahid, O, Rashid, I, Ali, Q, Akbar, MH, Oneeb, M, Shehzad, W, Ashraf, K, Sargison, N & Chaudhry, U 2019 'Genetic diversity and multiplicity of infection in *Fasciola gigantica* isolates of Pakistani livestock' bioRxiv, at Cold Spring Harbor Laboratory. <https://doi.org/10.1101/789297>

Digital Object Identifier (DOI):

[10.1101/789297](https://doi.org/10.1101/789297)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Genetic diversity and multiplicity of infection in *Fasciola gigantica* isolates of Pakistani livestock

Zia Ur Rehman^{2‡}, Osama Zahid^{1‡}, Imran Rashid², Qasim Ali³, Muhammad Haroon Akbar², Muhammad Oneeb², Wasim Shehzad⁴, Kamran Ashraf², Neil D. Sargison^{1*}, Umer Chaudhry^{1*}

1. *University of Edinburgh, Royal (Dick) School of Veterinary Studies and Roslin Institute, Easter Bush Veterinary Centre, Midlothian, Scotland, EH25 9RG, UK*

2. *Department of Parasitology, University of Veterinary and Animal Sciences Lahore, Punjab, Pakistan*

3. *Department of Parasitology, Gomal University Dera Ismail Khan, Khyber Pakhtoon Khaw, Pakistan*

4. *Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences Punjab, Pakistan*

‡ Contributed equally

*Corresponding authors: Umer Chaudhry

Email: uchaudhr@exseed.ed.ac.uk

Neil Sargison

Neil.Sargison@ed.ac.uk

Abstract

Fasciola gigantica liver flukes are responsible for over 3 billion US dollars of production loss annually in farmed livestock and cause widespread zoonotic disease. Nevertheless, the understating of the emergence and spread of the trematode species is poor. The multiplicity of *F. gigantica* infection and its spread is potentially influenced by multiple factors, including the abundance of suitable intermediate hosts, climatic conditions favoring the completion of the parasite's lifecycle, and translocation of infected animals or free-living parasite stages between regions. Here we describe the development of a 'tremabiome' metabarcoding sequencing method to explore the numbers of *F. gigantica* genotypes per infection and patterns of parasite spread, based on genetic characteristics of the mitochondrial NADH dehydrogenase 1 (mt-ND-1) locus. We collected *F. gigantica* from three abattoirs in the Punjab and Balochistan provinces of Pakistan, and our results show a high level of genetic diversity in 20 *F. gigantica* populations derived from small and large ruminants consigned to slaughter in both provinces. This implies that *F. gigantica* can reproduce in its definitive hosts through meiosis involving cross- and self-breeding, as described in the closely related species, *Fasciola hepatica*. The genetic diversity between the 20 populations derived from different locations also illustrates the impact of animal movements on gene flow. Our results demonstrate the predominance of single haplotypes, consistent with a single introduction of *F. gigantica* infection in 85% of the hosts from which the parasite populations were derived. This is consistent with clonal reproduction in the intermediate snail hosts.

Keywords: *Fasciola gigantica*, mitochondrial NADH dehydrogenase 1, gene flow, emergence of infection

1. Introduction

Fascioliosis and fascioliasis are important neglected worldwide diseases of ruminant livestock and humans, respectively. *Fasciola* spp. not only cause billion of dollar production losses annually in farm animals (Aghayan et al., 2019; Mungube et al., 2006), but are also emerging food-borne zoonoses in humans according to the World Health Organization (WHO), with 2.4 million people infected and over 180 million at risk (Haseeb et al., 2002; Henok and Mekonnen, 2011). The genus *Fasciola* comprises of two important species, *Fasciola gigantica* and *Fasciola hepatica* (Mas-Coma et al., 2005). *F. gigantica* typically cause disease in tropical regions (Amor et al., 2011), whereas *F. hepatica* is present in temperate areas (Mazeri et al., 2017). Overlap between these two species has been seen in subtropical regions, along with the presence of intermediate forms (Ichikawa and Itagaki, 2010; Rokni et al., 2010).

Animals become infected with *Fasciola* spp. following the ingestion of herbage contaminated with infective metacercariae, and the parasites' life cycles involve different snail species of *Galba truncatula* of Lymnaeidae family as intermediate hosts. Acquired parasite larvae penetrate the intestinal wall and migrate to the bile ducts of the definitive host, damaging the hepatic parenchyma and biliary tract (Geadkaew et al., 2011; Usip et al., 2014). Several potential factors may influence the emergence of *F. gigantica* infection. First, the seasonal presence of suitable intermediate hosts in areas with high rainfall, moderate temperatures, and humidity, and poor drainage (Claxton et al., 1997; Kaplan, 2001; Malek, 2018; Portugaliza et al., 2019; Rana et al., 2014; Rowcliffe and Ollerenshaw, 1960); resulting in high prevalences during certain months of the year (Phiri et al., 2005b; Rangel-Ruiz et al., 1999). Second, management stress, host immunity, and co-infection with intestinal nematodes; determining infection rates of *F. gigantica* (Ahmad et al., 2017; Elelu and Eisler, 2018; Phiri et al., 2005a). Third, the clonal expansion of *Fasciola* within its intermediate snail host; resulting in pasture contamination and subsequent host infection by metacercariae of the same genetic origin. This clonal expansion can also produce a genetic bottleneck effect when levels of infection in the snail populations are low (Beesley et al., 2017).

Over past few decades, high levels of animal movement have been reported in domestic ruminants in several European and Asian countries (Kelley et al., 2016; Vilas et al., 2012); hence genetic analyses are needed to understand the corresponding spread of *F. gigantica* infections and aid in the development of parasite control strategies (Hayashi et al., 2016).

Animal movement patterns differ between farms and *F. gigantica* infects a wide range of hosts including farm and wild animals and humans. This potentially enables wildlife reservoirs of infection and cross-infection between different host species, complicating control strategies (Rojo-Vázquez et al., 2012).

Genetic relationships between parasite species and populations can be analysed using internal transcribed spacer region 2 (rDNA ITS-2) and mitochondrial nicotinamide adenine dinucleotide (NADH) dehydrogenase 1 (mt-ND-1) sequence data, respectively. Metabarcoding and deep amplicon sequencing methods and the Illumina Mi-Seq platform enable the study of large and diverse fluke populations, to show if infection has emerged recently in the host at a single point of time, or if burdens have been established repeatedly at different times, before spreading as a result of animal movement. Recently, we have used these methods to study the multiplicity of *Calicophoron daubneyi* infection in United Kingdom cattle herds. Our findings were consistent with multiple independent emergences of *C. daubneyi* infection; while the identification of common mt-COX-1 haplotypes in several populations across a range of geographic locations, highlighting the role of animal movements in the parasite's spread (Sargison et al., 2019).

In this paper we describe a study using liver fluke parasites collected from cattle, buffalo, sheep and goats slaughtered in three geographically separated abattoirs in the Punjab and Balochistan provinces of Pakistan, with the following aims: i) to confirm the species identity of recovered *Fasciola* spp.; ii) to identify the presence of single or multiple genotypes per infection (multiplicity of infection) and iii) demonstrate the spread of *F. gigantica* mt-ND-1 haplotypes. The species identity of liver flukes was confirmed by deep amplicon sequencing of 483bp metabarcoding fragments of the rDNA ITS-2 region. Haplotype diversity in 20 *F. gigantica* populations, each derived from single infected ruminants consigned from different locations, was shown by deep amplicon sequencing of 311bp metabarcoding fragments of the mt-ND-1 locus. Split and network trees of the mt-ND-1 haplotypes were examined to show the multiplicity of *F. gigantica* infection and the patterns their spread, providing proof of concept for a novel approach to epidemiological studies of fasciolosis and fascioliasis, and validation of parasite control strategies.

2. Materials and methods

2.1. Sample collection and genomic DNA extraction

Three abattoirs were selected in the Punjab and Balochistan provinces of Pakistan, where there is a known high prevalence of fasciolosis. A total of 26 infected livers were collected from the three city abattoirs. Fourteen samples (buffalo=7, cattle=1, goat=6) were collected from Lahore, Punjab (31.5204° N, 74.3587° E), six samples (cattle=4, sheep=2) were collected from Murgha Kibzai, Balochistan (30.7384° N, 69.4136° E), and six samples (buffalo=3, sheep=3) were obtained from Loralai, Balochistan (30.3806° N, 68.5963° E). The livers were transported from abattoir to the laboratory on ice, where liver flukes were extracted from the biliary ducts by dissection. In total, 305 (183 from Punjab and 122 from Balochistan) individual flukes obtained from the populations. The flukes were identified as *Fasciola* spp. morphologically, thoroughly washed with phosphate-buffered saline (PBS), and preserved in 70% ethanol at -80 °C. For DNA extraction, a small tissue piece of around 2 mg was taken from the head of each worm, avoiding the possible contamination with any eggs. Each piece was rinsed twice for 5 min each, in a petri dish with distilled water (dH₂O) and then lysed in 25 µl worm-lysis solution, prepared by mixing of 50 µl Proteinase K (10 mg/ml, New England BioLabs) and 50 µl 1M Dichloro Diphenyl Trichloroethane (DDT) in 1 ml of DirectPCR Lysis Reagent (Viagen). The lysates were then incubated at 60°C for 2 hrs, followed by 85°C for 15 min. Lysates were stored at -80°C until use. The abattoir-based study was approved by the Ethical Review Committee of the University of Veterinary and Animal Sciences Lahore Pakistan.

2.2. Metabarcoded sequencing approach

The overall scheme of the metabarcoding sequencing approach using Illumina MiSeq platform is shown in Fig. 1. The whole procedure was run in twice: first on the individual worms for species identification using the rDNA ITS-2 marker; then on pooled *F. gigantica* from each population, using the mt-ND-1 marker.

2.2.1. Adapter PCR amplification of rDNA ITS-2 and mt-ND-1 loci

The 1st round PCR amplification was performed on 483bp fragments of the rDNA ITS-2 region, complementary to the 5.8s and 28s rDNA coding sequences, using universal sets of primers (Adlard et al., 1993; Chaudhry et al., 2016) and 311bp mt-ND-1 fragments, using newly developed primers (Supplementary Table S1, Fig. 1). The metabarcoding adapters were

added to the primers to allow their successive annealing, and the enhancement of the amplicons by adding up to three random nucleotides (N) between each primer and adapter, resulting in a total of four forward and four reverse primers (Supplementary Table S1). For both rDNA ITS-2 and mt-ND-1 loci, equal proportions of four forward and four reverse primers were mixed and used for the adapter PCR with the following conditions: 10 mM dNTPs, 10 μ M forward and reverse adapter primers, 5X KAPA HiFi Fidelity buffer, 0.5 U KAPA HiFi Fidelity Polymerase (KAPA Biosystems, USA), 14 μ l ddH₂O and 1 μ l of fluke lysate. The thermocycling conditions of the PCR reaction were as follows: 95°C for 2 min, followed by 35 cycles of 98°C for 20 sec, 65°C for 15 sec for ITS-2, and 60°C for 15 sec for ND-1, and 72°C for 15 sec, followed by a final extension at 72°C for 5 min. PCR products were then purified with AMPure XP Magnetic Beads (1X) according to the manufacturer's instructions (Beckman Coulter).

2.2.2. Barcoded PCR amplification of rDNA ITS-2 and mt-ND-1 loci

In the 2nd round metabarcoded PCR, 16 forward (N501-N516) and 24 reverse (N701-N724) primers were used (Supplementary Table S2, Fig. 1) in a way that each sample contained a unique combination of forward and reverse primers. The barcoded PCR was performed with the following conditions: 10 mM dNTPs, 10 μ M barcoded forward and reverse primers, 5X KAPA HiFi Fidelity buffer, 0.5 U KAPA HiFi Fidelity Polymerase (KAPA Biosystems, USA), 14 μ l ddH₂O and 2 μ l of the first-round PCR product. The thermocycling conditions for this round of PCR were: 98°C for 45 sec, followed by 7 cycles of 98°C for 20 sec and 63°C for 20 sec, and final extension at 72°C for 2 min. The PCR products were purified with AMPure XP Magnetic Beads (1X) according to the manufacturer's instructions (Beckman Coulter).

2.2.3. Illumina Mi-Seq run, bioinformatics data handling and analysis

The bead-purified products from each sample were mixed to prepare a pooled library (Fig. 1) and measured with the KAPA qPCR library quantification kit (KAPA Biosystems, USA) before running it on the Illumina Mi-Seq Sequencer using a 500-cycle pair-end reagent kit (Mi-Seq Reagent Kits v2, MS-103-2003) at a concentration of 15nM with addition of 15% Phix Control v3 (Illumina, FC-11-2003). During the post-run processing, Mi-Seq splits all sequences by samples using the barcoded indices to produce FASTQ files (Supplementary

Table S2). The analysis of both rDNA ITS-2 and mt-ND-1 FASTQ files were performed in Mothur v1.39.5 software (Schloss et al., 2009), using our modified Command Prompt pipeline (Sargison et al., 2019), and the standard operating procedures of Illumina Mi-Seq (Kozich et al., 2013).

For both rDNA ITS-2 and mt-ND-1, the raw paired-end reads were analysed to combine the two sets of reads for each parasite population using `make.contigs` command, which requires ‘`stability.files`’ as an input. The ‘`make.contigs`’ command extracts sequence and quality score data from FASTQ files, creating complements of the reverse and forward reads and joins them into contigs. It simply aligns the pairs of sequence reads and compares the alignments to identify any positions where the two reads disagree. Next, there was a need to remove any sequences with ambiguous bases using the ‘`screen.seqs`’ command. The alignment of the above dataset was performed with a *F. gigantica* and *F. hepatica* rDNA ITS-2 and *F. gigantica* mt-ND-1 reference sequence taxonomy library created from the NCBI database (for more details Supplementary Data S1; Section 2.4) based on where the sequences start and end corresponding with the primer set (Supplementary Table S1). The ‘`align.seqs`’ command was adapted for the reference sequence taxonomy library to be aligned with rDNA ITS-2 and mt-ND-1 Illumina Mi-Seq dataset. To confirm that these filtered sequences overlap the same region of the reference sequence taxonomy library, the ‘`screen.seqs`’ command was run to show the sequences ending at the 483bp rDNA ITS-2 and 311bp mt-ND-1 position.

At this stage, the rDNA ITS-2 analysis was completed by classifying the sequences into either of the two groups (*F. gigantica* or *F. hepatica*), using the `classify.seqs` command and creating the taxonomy file by using the `summary.tax` command. Overall, thousands of rDNA ITS-2 reads were generated from the data set of 305 individual worms from 26 populations. The presence of each species was calculated by dividing the number of sequences reads of each worm by the total number of reads.

For mt-ND-1 analysis, once all sequences were classified as *F. gigantica* mt-ND-1 based on the ‘`screen.seqs`’ command, a count list of the consensus sequences of each population was created using the `unique.seqs` command, followed by the use of `pre.cluster` command to look for sequences having up to two differences and to merge them in groups based on their abundance. Chimeras were identified and removed using the `chimera.vsearch` commands. The count list was then used to create the FASTQ files of the consensus sequences of each population using the `split.groups` command (for more details Supplementary Data S2). The

consensus sequences of *F. gigantica* mt-ND-1 were analysed separately in Geneious v9.0.1 software (Biomatters Ltd, New Zealand) (Kearse et al., 2012) using the MUSCLE alignment tool to remove the polymorphisms occurring once only as being artifacts due to sequencing errors as previously described (Sargison et al., 2019). These aligned consensus sequences were then imported into the FaBox 1.5 online tool to collapse the sequences that showed 100% base pair similarity after corrections into a single haplotype. The frequency of all the haplotypes present in the total and at each population level was calculated by dividing the number of sequences reads of each population by the total number of reads.

2.3. Network and Split tree analysis

A network tree was produced based on a neighbor-joining algorithm using Network 4.6.1 software (Fluxus Technology Ltd), built on a sparse network with the epsilon parameter is set to zero default. A split tree was created in the SplitTrees4 software (Huson and Bryant, 2005) by using the UPGMA method in the Jukes-Cantor model of substitution. The appropriate model of nucleotide substitutions for UPGMA analysis was selected by using the jModeltest 12.2.0 program (Posada, 2008). A maximum-likelihood tree of was also constructed in Mega 6.0 (Tamura et al., 2013) based on a neighbor-joining algorithm using HKY+G+I model. The branch supports were obtained by 1000 bootstraps of the data.

Genetic diversity was calculated for all the generated sequences within and between populations by using the DnaSP 5.10 software package (Librado and Rozas, 2009), and the following values were obtained: Haplotype diversity (H_d), the number of segregating sites (S), nucleotide diversity (π), the mean number of pairwise differences (k), the mutation parameter based on an infinite site equilibrium model, and the number of segregating sites (θ_S).

3. Results

3.1. Consensus sequence taxonomy library preparation of *F. gigantica* and *F. hepatica* rDNA ITS-2 and mt-ND-1 loci

A total of 27 NCBI GeneBank sequences of the *Fasciola* rDNA ITS-2 locus (*F. gigantica*=14, *F. hepatica*=13) and 167 NCBI GeneBank sequences of the *Fasciola* mt-ND-1 locus (*F. gigantica*=67, *F. hepatica*=100) were identified by BLAST search (for more details

Supplementary Data S1). The sequences were analysed separately in Geneious v9.0.1 software (Biomatters Ltd, New Zealand) (Kearse et al., 2012) using the MUSCLE alignment tool. The construction of the maximum-likelihood tree demonstrate distinct clustering of rDNA ITS-2 and mt-ND-1 loci of *F. gigantea* and *F. hepatica* (Fig. 2A & B).

3.2. Confirmation of species identity and co-infection by rDNA ITS-2 sequence analysis

A total 305 individual worms, comprising of 26 populations, were run through the Illumina MiSeq platform, out of which 238 (77.6%) were confirmed to be *F. gigantea*, 56 (18.7%) were *F. hepatica* and 11 (3.7%) showed heterozygous sequence reads (Table 1). All 183 worms collected in Punjab from buffalo, cattle, and goats were *F. gigantea*. In contrast, 122 worms collected from Balochistan were both *F. gigantea* (42.2%) and *F. hepatica* (48.3%), as well as the intermediate form (9.5%); except the buffalo populations contained only *F. gigantea* (Table 1).

3.3. Haplotype distribution of *F. gigantea* mt-ND-1 locus

The haplotype frequencies present at the individual population level were analysed separately from 20 selected *F. gigantea* populations (Fig. 3). Only three populations (P8B, P12G, and P18C) each containing between 7 and 8 worms, showed relatively high frequencies of multiple haplotypes. These three populations contained 7, 3, and 13 haplotypes (Fig. 3). In contrast, single haplotypes predominated in 17 populations, each containing between 4 and 20 worms (Fig. 3). Three of these populations (P13B, P7B, and P17C) contained 6 haplotypes; six populations (P2B, P14B, P15B, P16B, P19B, and P26C) had 5 haplotypes; three populations (P4B, P6G, and P10G) had 4 haplotypes; four populations (P1G, P9G, P3B, and P11G) had 3 haplotypes; and one population (P5B) contained only one haplotype (Fig. 3).

Twenty-six haplotypes of the *F. gigantea* mt-ND-1 locus were identified among all 20 populations. The split tree analysis shows at least two distinct clades (Fig. 4). Six haplotypes (H13, H26, H4, H19, H9, and H20) in Clade I, were present only in Balochistan. The remaining 20 haplotypes were in clade II, with 8 haplotypes (H17, H21, H14, H10, H12, H5, H7, and H6) predominating in Balochistan. Contrastingly, the six haplotypes (H17, H21, H14, H10, H12, H5, H7, and H6) were more common in Punjab. In addition, four equally shared haplotypes (H2, H3, H11, and H23) were present in clade II.

3.4. Phylogenetic analysis of *F. gigantea* mt-ND-1 locus

The network and maximum likelihood trees were produced to examine the phylogenetic relationship between 26 different haplotypes of the *F. gigantea* mt-ND-1 locus (Fig. 5, Supplementary Fig. S1). Sixteen haplotypes were comparatively rare as they collectively made less than 5% of the total reads. Five of these haplotypes (H15, H16, H18, H22, and H25) were only found in Punjab, and they were present in four populations; whereas 11 haplotypes (H5, H7, H9, H12, H13, H14, H17, H19, H20, H21, and H26) were present in Balochistan, originating from 5 populations (Fig. 5, Supplementary Fig. S1). The remaining 10 haplotypes were shared between Punjab and Balochistan, making more than 95% of the total reads. Three of these haplotypes (H1, H8, and H24) predominated in Punjab and were present in 13 populations in total, including 9 populations from Punjab and 4 from Balochistan (Fig. 5, Supplementary Fig. S1). Three haplotypes (H4, H6, and H10) predominated in Balochistan and were present in 14 populations in total, including 5 from Balochistan and 9 from Punjab (Fig. 5, Supplementary Fig. S1). In contrast, four haplotypes (H2, H3, H11, and H23) were equally shared between Punjab and Balochistan (Fig. 5, Supplementary Fig. S1); these were present in a total of 19 populations, 13 coming from Punjab and 6 from Balochistan.

3.5. Genetic diversity of *F. gigantea* mt-ND-1 locus

The data show a high level of genetic diversity at both haplotype and nucleotide levels (Table 2), with the values of haplotype diversity ranging from 0.727 to 0.960 and nucleotide diversity from 0.00643 to 0.03118 within different populations. The overall values between populations were 0.917 and 0.02466 for haplotype and nucleotide diversity, respectively. This analysis confirms that Pakistani *F. gigantea* populations have a high level of genetic diversity.

4. Discussion

Next-generation genomic resources have potential applications to investigate parasite species dynamics, coinfections, hybridisation, multiplicity of infection and the level of gene flow, as well as evaluation of population responses to drug treatments (Sargison et al., 2019). Various PCR methods [(RFLP)-PCR, quantitative (q)PCR, multiplex PCR, SSCP-PCR,

Sanger sequencing] have been described to amplify the mt-ND-1 and rDNA ITS-2 regions for sequence determination (Ai et al., 2011), but these are low throughput, hence relatively expensive, and potentially error-prone. In contrast, high throughput deep amplicon sequencing or metabarcoded DNA derived from parasite populations using the Illumina MiSeq platform is relatively low-cost and potentially less error-prone. The method has transformed the study of haemoprotozoan parasites (Chaudhry et al., 2019) and nematode parasites (Avramenko et al., 2015); and has the potential to open new areas of research in the study of trematode parasite communities. We have used this method to study the multiplicity of *C. daubneyi* infection in the United Kingdom (Sargison et al., 2019). The use of primers binding to conserved sites and analysis of up to 600 bp sequence reads allows trematode species (the ‘tremabiome’) to be determined. The use of barcoded primers allows a large number of samples to be pooled and sequenced in a single Mi-Seq run, making the technology suitable for high-throughput analysis. By multiplexing the barcoded primer combinations, it is possible to run 384 samples at once on a single Illumina Mi-Seq flow cell, helping to reduce the cost (Chaudhry et al., 2019; Sargison et al., 2019). In the present study, using high throughput deep amplicon sequencing of metabarcoded DNA, we first confirm the presence of *F. gigantica* and then identify the prevalence of single or multiple haplotypes per infection (multiplicity of infection) and demonstrating the spread of *F. gigantica* alleles.

Fasciola spp. are cosmopolitan parasites that are widely distributed throughout the world (Mas-Coma, 2004). The prevalence of *F. gigantica* is highest in tropical and subtropical regions while *F. hepatica* is more common in temperate areas where sheep and cattle are raised (Mas-Coma et al., 2014). Pakistan has an agriculture-based economy with livestock being an integral part. There are several reports of *Fasciola* infections in Pakistan, mostly implying that *F. hepatica* is most commonly identified in small and large ruminants (Ashraf et al., 2014; Ijaz et al., 2009; Shahzad et al., 2012). A few reports indicate the presence of both *F. hepatica* and *F. gigantica*, with intermediate forms found (Akhtar et al., 2012; Iqbal et al., 2007). However, all of these reports are based on egg and adult morphology with limited molecular confirmation of species identity. In the present study of the dynamics of *Fasciola* spp. in Pakistan, ITS-2 rDNA sequence data confirmed the presence of *F. hepatica*, *F. gigantica*, and intermediate forms, with *F. gigantica* being the predominant species (Table 1), concurring with a previous study (Chaudhry et al., 2016).

The predominance of single mt-ND-1 haplotypes in 17 of 20 (85%) *F. gigantica* populations suggests a single genetic emergence of infection in the host from which the parasites were derived. In contrast, the presence of high proportions of multiple mt-ND-1 haplotypes in just 3 of 20 (15%) populations implies multiple emergences (Fig. 3). The clonal lineage of the haplotypes in most of the *F. gigantica* populations is similar to the situation that has been described in *F. hepatica* (Beesley et al., 2017; Vilas et al., 2012), implying similar importance of asexual reproduction of the intermediate snail hosts to the population genetics of the two species. The clonal lineage might arise because the life cycle of *F. gigantica* involves uneven disease transmission intensity and, or, abundance of intermediate hosts; as has been described for *F. hepatica*, where single miracidia infect the *G. truncatula* mud snails, giving rise to multiple, genetically identical cercariae (Beesley et al., 2017). Then, after the release of clonal lineage cercariae by the intermediate hosts, there may be an aggregation of infective metacercariae, with little mixing on the pasture, prior to the ingestion by the definitive host. These conditions would be more likely if the prevalence of infection in the intermediate hosts is low, and snail habitats are small and isolated, as is likely the case in Pakistan, where long-term muddy regions tend to be found in association with irrigation pumps or ponds. An alternative explanation for the clonal lineage of the *F. gigantica* could be a little population bottlenecking arising due to the influences of the wide range of climatic conditions throughout the year on intermediate snail host infection (Rana et al., 2014).

Our results confirm a high level of genetic diversity in *F. gigantica* (Table 2), implying the capacity for reproduction in the definitive host through meiosis during cross-breeding. A similar situation has been described in *F. hepatica* infection (Beesley et al., 2017; Vilas et al., 2012; Zintl et al., 2015). However, a low level of infection (Dar et al., 2011; Rondelaud et al., 2011) and clonal lineage in snails could give rise to a potential low population bottlenecking effect, raising questions about how genetic diversity might be maintained in *F. gigantica*. For example, could snails be infected with multiple miracidia and subsequently shed cercariae with different alleles, or could diversity be maintained by long survival of metacercariae on the pasture?.

The movement of livestock is frequent in Pakistan and contributes to the high levels of parasitic gene flow (Ali et al., 2018). In the present study, ten haplotypes were spread between populations. Three of those haplotypes predominated in the Punjab and three in the Balochistan provinces. In contrast, four haplotypes were equally shared between the

populations from both provinces. This suggests that animal movement plays a key role in maintaining a high level of gene flow in *F. gigantica*, for example translocating animals to a new region could introduce new parasites as well as exposing the definitive hosts to different residents parasite populations.

In summary, we used a mt-ND-1 marker to provide first insights into the genetic diversity and multiplicity of *F. gigantica* infection in Pakistan. Our findings suggest that most of the hosts were predominantly infected with parasites of identical mt-ND-1 haplotypes, consistent with clonal multiplication within the snails. A high level of genetic diversity was seen in *F. gigantica* isolated from naturally infected small and large ruminants, showing that sexual reproduction with cross-fertilisation occurs within the definitive host. The most common mt-ND-1 haplotypes were identified in several populations across a range of geographic locations, highlighting the role of animal movements in the spread of infectious disease. Our study provides proof of concept for a method that could be used to investigate host-parasite relationships and to examine influences of intermediate hosts, co-infections and climate change on the epidemiology of *F. gigantica*; with reference to the development of sustainable parasite control strategies.

Acknowledgment

The study was financially supported by Roslin Institute using facilities funded by Biotechnology and Biological Sciences Research Council (BBSRC). Work at the University of Veterinary and Animal Science Pakistan uses facilities funded by the Higher Education Commission of Pakistan. The authors of this study would like to thank Dr. Talat Naseer Pasha - Vice-Chancellor of the University of Veterinary and Animal Science Lahore Pakistan for his great support in the arrangements for sample collections.

Conflict of interest

None

References

- Adlard, R.D., Barker, S.C., Blair, D., Cribb, T.H., 1993. Comparison of the second internal transcribed spacer (ribosomal DNA) from populations and species of Fasciolidae (Digenea). *International journal for parasitology* 23, 423-425.
- Aghayan, S., Gevorgian, H., Ebi, D., Atoyan, H.A., Addy, F., Mackenstedt, U., Romig, T., Wassermann, M., 2019. *Fasciola* spp. in Armenia: Genetic diversity in a global context. *Veterinary parasitology* 268, 21-31.
- Ahmad, I., Durrani, A., Ashraf, K., Hameed, K., Khan, O., Abbasi, A., Shakar, A., Hameed, N., Haq, I., Safdar, A., 2017. Risk factors analysis of fasciolosis in bufflo in azad jammu and kashmir. *Pakistan Journal of Science* 69.
- Ai, L., Chen, M.X., Alasaad, S., Elsheikha, H.M., Li, J., Li, H.L., Lin, R.Q., Zou, F.C., Zhu, X.Q., Chen, J.X., 2011. Genetic characterization, species differentiation and detection of *Fasciola* spp. by molecular approaches. *Parasites & vectors* 4, 101.
- Akhtar, A., Arshad, M., Ameer, M., 2012. Prevalence of *Fasciola hepatica* in sheep and goats in district Dera Ismail Khan. *Pakistan Journal of Science* 64, 31.
- Ali, Q., Rashid, I., Shabbir, M.Z., Shahzad, K., Ashraf, K., Sargison, N.D., Chaudhry, U., 2018. Population genetics of benzimidazole-resistant *Haemonchus contortus* and *Haemonchus placei* from buffalo and cattle: implications for the emergence and spread of resistance mutations. *Parasitology research* 117, 3575-3583.
- Amor, N., Farjallah, S., Salem, M., Lamine, D.M., Merella, P., Said, K., Slimane, B.B., 2011. Molecular characterization of *Fasciola gigantica* from Mauritania based on mitochondrial and nuclear ribosomal DNA sequences. *Experimental parasitology* 129, 127-136.
- Ashraf, S., Iqbal, Z., Ali, M., Chaudhry, H.R., Sial, N., Ahsan, U., Ali, A., Asif, M.Z., 2014. Seasonal prevalence of *Fasciola hepatica* infection in buffaloes of Bahawalpur district of Punjab, Pakistan. *Journal of Infection and Molecular Biology* 2, 30-31.
- Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D., Gilleard, J.S., 2015. Exploring the gastrointestinal “nemabiome”: deep amplicon sequencing to quantify the species composition of parasitic nematode communities. *PLoS One* 10, e0143559.
- Beesley, N.J., Williams, D.J., Paterson, S., Hodgkinson, J., 2017. *Fasciola hepatica* demonstrates high levels of genetic diversity, a lack of population structure and high gene flow: possible implications for drug resistance. *International journal for parasitology* 47, 11-20.
- Chaudhry, U., Ali, Q., Rashid, I., Shabbir, M.Z., Ijaz, M., Abbas, M., Evans, M., Ashraf, K., Morrison, I., Morrison, L., Sargison, N.D., 2019. Development of a deep amplicon sequencing method to determine the species composition of piroplasm haemoprotozoa. *Ticks Tick Borne Dis* 10, 101276.
- Chaudhry, U., Van Paridon, B., Shabbir, M., Shafee, M., Ashraf, K., Yaqub, T., Gilleard, J., 2016. Molecular evidence shows that the liver fluke *Fasciola gigantica* is the predominant *Fasciola* species in ruminants from Pakistan. *Journal of helminthology* 90, 206-213.
- Claxton, J.R., Zambrano, H., Ortiz, P., Amorós, C., Delgado, E., Ecurra, E., Clarkson, M.J., 1997. The epidemiology of fasciolosis in the inter-Andean valley of Cajamarca, Peru. *Parasitology international* 46, 281-288.
- Dar, Y., Vignoles, P., Dreyfuss, G., Rondelaud, D., 2011. The development of rediae of *Fasciola hepatica* in *Radix natalensis* subjected twice to bimiracidial exposures. *Journal of helminthology* 85, 210-214.
- Elelu, N., Eisler, M., 2018. A review of bovine fasciolosis and other trematode infections in Nigeria. *Journal of helminthology* 92, 128-141.
- Geadkaew, A., Von Bülow, J., Beitz, E., Grams, S.V., Viyanant, V., Grams, R., 2011. Functional analysis of novel aquaporins from *Fasciola gigantica*. *Molecular and biochemical parasitology* 175, 144-153.
- Haseeb, A., El-Shazly, A., Arafa, M., Morsy, A., 2002. A review on fascioliasis in Egypt. *Journal of the Egyptian Society of Parasitology* 32, 317-354.
- Hayashi, K., Ichikawa-Seki, M., Allamanda, P., Wibowo, P.E., Mohanta, U.K., Sodorun, Guswanto, A., Nishikawa, Y., 2016. Molecular characterization and phylogenetic analysis of *Fasciola gigantica* from western Java, Indonesia. *Parasitol Int* 65, 424-427.

- Henok, M., Mekonnen, A., 2011. Study on the prevalence and risk factors of Fasciolosis in small ruminants in and around Hirna Town, Ethiopia. *Global Veterinaria* 7, 497-501.
- Huson, D.H., Bryant, D., 2005. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* 23, 254-267.
- Ichikawa, M., Itagaki, T., 2010. Discrimination of the ITS1 types of *Fasciola* spp. based on a PCR-RFLP method. *Parasitology research* 106, 757-761.
- Ijaz, M., Khan, M., Avais, M., Ashraf, K., Ali, M., Khan, M., 2009. Infection rate and chemotherapy of various helminthes in diarrhoeic sheep in and around Lahore. *J Anim Plant Sci* 19, 13-16.
- Iqbal, M., Sajid, M., Hussain, A., Khan, M., 2007. Prevalence of helminth infections in dairy animals of Nestle milk collection areas of Punjab (Pakistan). *Italian Journal of Animal Science* 6, 935-938.
- Kaplan, R.M., 2001. *Fasciola hepatica*: a review of the economic impact in cattle and considerations for control. *Vet Ther* 2, 40-50.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649.
- Kelley, J.M., Elliott, T.P., Beddoe, T., Anderson, G., Skuce, P., Spithill, T.W., 2016. Current threat of triclabendazole resistance in *Fasciola hepatica*. *Trends in parasitology* 32, 458-469.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112-5120.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-1452.
- Malek, E.A., 2018. *Snail Transmitted Parasitic Diseases: Volume II*. CRC Press.
- Mas-Coma, S., 2004. Human fascioliasis: epidemiological patterns in human endemic areas of South America, Africa and Asia. *Southeast Asian J Trop Med Public Health* 35, 1-11.
- Mas-Coma, S., Bargues, M.D., Valero, M., 2005. Fascioliasis and other plant-borne trematode zoonoses. *International journal for parasitology* 35, 1255-1278.
- Mas-Coma, S., Valero, M.A., Bargues, M.D., 2014. Fascioliasis. *Advances in experimental medicine and biology* 766, 77-114.
- Mazeri, S., Rydevik, G., Handel, I., Bronsvoort, B.M.D., Sargison, N., 2017. Estimation of the impact of *Fasciola hepatica* infection on time taken for UK beef cattle to reach slaughter weight. *Scientific reports* 7, 7319.
- Mungube, E., Bauni, S., Tenhagen, B.-A., Wamae, L., Nginyi, J., Mugambi, J., 2006. The prevalence and economic significance of *Fasciola gigantica* and *Stilesia hepatica* in slaughtered animals in the semi-arid coastal Kenya. *Tropical Animal Health and Production* 38, 475-483.
- Phiri, A., Phiri, I., Sikasunge, C., Monrad, J., 2005a. Prevalence of fasciolosis in Zambian cattle observed at selected abattoirs with emphasis on age, sex and origin. *Journal of veterinary medicine, series B* 52, 414-416.
- Phiri, A., Phiri, I., Siziya, S., Sikasunge, C., Chembensofu, M., Monrad, J., 2005b. Seasonal pattern of bovine fasciolosis in the Kafue and Zambezi catchment areas of Zambia. *Veterinary parasitology* 134, 87-92.
- Portugaliza, H.P., Balaso, I.M.C., Descallar, J.C.B., Lañada, E.B., 2019. Prevalence, risk factors, and spatial distribution of *Fasciola* in carabao and intermediate host in Baybay, Leyte, Philippines. *Veterinary Parasitology: Regional Studies and Reports* 15, 100261.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* 25, 1253-1256.
- Rana, M., Roohi, N., Khan, M., 2014. Fascioliasis in cattle-a review. *J. Anim. Plant Sci* 24, 668-675.
- Rangel-Ruiz, L., Marquez-Izquierdo, R., Bravo-Nogueira, G., 1999. Bovine fasciolosis in Tabasco, Mexico. *Veterinary parasitology* 81, 119-127.
- Rojo-Vázquez, F.A., Meana, A., Valcárcel, F., Martínez-Valladares, M., 2012. Update on trematode infections in sheep. *Veterinary parasitology* 189, 15-38.

- Rokni, M.B., Mirhendi, H., Mizani, A., Mohebbi, M., Sharbatkhori, M., Kia, E.B., Abdoli, H., Izadi, S., 2010. Identification and differentiation of *Fasciola hepatica* and *Fasciola gigantica* using a simple PCR-restriction enzyme method. *Experimental parasitology* 124, 209-213.
- Rondelaud, D., Hourdin, P., Vignoles, P., Dreyfuss, G., Cabaret, J., 2011. The detection of snail host habitats in liver fluke infected farms by use of plant indicators. *Veterinary parasitology* 181, 166-173.
- Rowcliffe, S., Ollerenshaw, C., 1960. Observations on the bionomics of the egg of *Fasciola hepatica*. *Annals of Tropical Medicine & Parasitology* 54, 172-181.
- Sargison, N.D., Shahzad, K., Mazeri, S., Chaudhry, U., 2019. A high throughput deep amplicon sequencing method to show the emergence and spread of *Calicophoron daubneyi* rumen fluke infection in United Kingdom cattle herds. *Veterinary parasitology*.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537-7541.
- Shahzad, W., Mehmood, K., Munir, R., Aslam, W., Ijaz, M., Ahmad, R., Khan, M.S., Sabir, A.J., 2012. Prevalence and molecular diagnosis of *Fasciola hepatica* in sheep and goats in different districts of Punjab, Pakistan. *Pak Vet J* 32, 535-538.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* 30, 2725-2729.
- Usip, L., Ibanga, E., Edoho, H., Amadi, E., Utah, E., 2014. Prevalence of Fascioliasis and the economic loss of condemned liver due to *Fasciola* infection in Cattle slaughtered at three abattoirs in Eket Urban, Akwa Ibom State of Nigeria. *Global Advanced Research Journal of Food Science and Technology* 3, 54-75.
- Vilas, R., Vázquez-Prieto, S., Paniagua, E., 2012. Contrasting patterns of population genetic structure of *Fasciola hepatica* from cattle and sheep: implications for the evolution of anthelmintic resistance. *Infection, Genetics and Evolution* 12, 45-52.
- Zintl, A., Talavera, S., Sacchi-Nestor, C., Ryan, M., Chrysafidis, A., Mulcahy, G., 2015. Comparison of *Fasciola hepatica* genotypes in relation to their ability to establish patent infections in the final host. *Veterinary parasitology* 210, 145-150.

Figure Legends

Fig. 1. Schematic representation of the preparation and bioinformatics analysis of the metabarcoding sequencing libraries. In the first-round PCR amplification, overhanging primers were used to amplify the ITS-2 rDNA and mt-ND-1 fragments. The adapter base pairs provide the target sites for the primers used for sequencing the fragment. The random nucleotides ('0-3N's) are inserted between the adapter and the primers to offset the reading frame, thereby amplicons were sequenced to prevent the oversaturation of the MiSeq sequencing channels. The second-round PCR amplification was then performed using overhanging barcoded primers to bind to the adapter tags to add indices as well as the P5 and P7 regions required to bind to the MiSeq flow cell. The analysis of both rDNA ITS-2 and mt-ND-1 FASTQ files were performed in Mothur v1.39.5 software (Schloss et al., 2009) by using our modified Command Prompt pipeline and the standard operating procedures of Illumina MiSeq (Kozich et al., 2013).

Fig. 2. The maximum-likelihood tree was obtained from the BLAST searched *F. gigantica* and *F. hepatica* rDNA ITS-2 and mt-ND-1 loci. The NCBI GeneBank sequences were first aligned on the MUSCLE tool of the Geneious v9.0.1 software. (A) Twenty-seven sequences of the rDNA ITS-2 were identified among *F. gigantica* and *F. hepatica* species. (B) Forty

unique sequences of the mt-ND-1 were identified among hundred and sixty-seven *F. gigantea* and *F. hepatica* species. The neighbor-joining algorithm (Kimura 2 and HKY+G+I-parameter model) was computed with 1000 bootstrap replicates using MEGA5 software created by Biomatters. Each species was identified with different shades bars.

Fig. 3. Relative allele frequencies of twenty *F. gigantea* populations from Punjab (A) and Balochistan (B) provinces of Pakistan. Each pie chart displays the individual population, and each haplotype in the pie chart is represented by a different color codes. The pie chart circle represents the population distribution and their frequency comes from each of the twenty-six haplotypes as indicated on the insert table. The populations are also linked to the location from where they were collected.

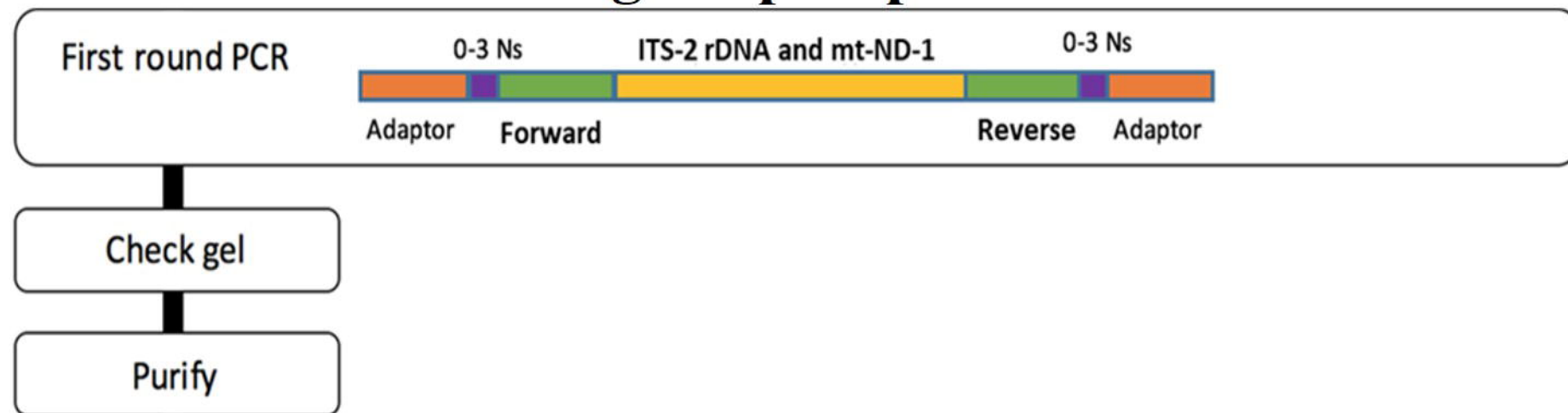
Fig. 4. Split tree of twenty-six mt-ND-1 haplotypes generated from twenty *F. gigantea* populations. The tree was constructed with the UPGMA method in the Jukes-Cantor model of substitution in the SplitsTrees4 software. The appropriate model of nucleotide substitutions for UPGMA analysis was selected by using the jModeltest 12.2.0 program. The pie chart circles in the tree represent the haplotypes in two different clades and the size of each circle is proportional to the frequency in the population.

Fig. 5. Network tree of twenty-six mt-ND-1 haplotypes sequenced from twenty *F. gigantea* populations. The tree was generated with the Neighbour Joining method in the Network 4.6.1 software (Fluxus Technology Ltd.). All unnecessary median vectors and links were removed with the star contractions. Each pie chart displays the individual haplotype, and each population in the pie chart is represented by a different color codes. The pie chart circle represent the haplotype distribution and their frequency comes from each of the twenty populations as indicated on the insert table. The color of each haplotype circle represents the percentage of sequence reads generated per population shown in the insert table. The number of mutations (in red) separating adjacent sequence nodes was indicated along with the connecting branches and the length of the lines connecting the haplotypes is proportional to the number of nucleotide changes.

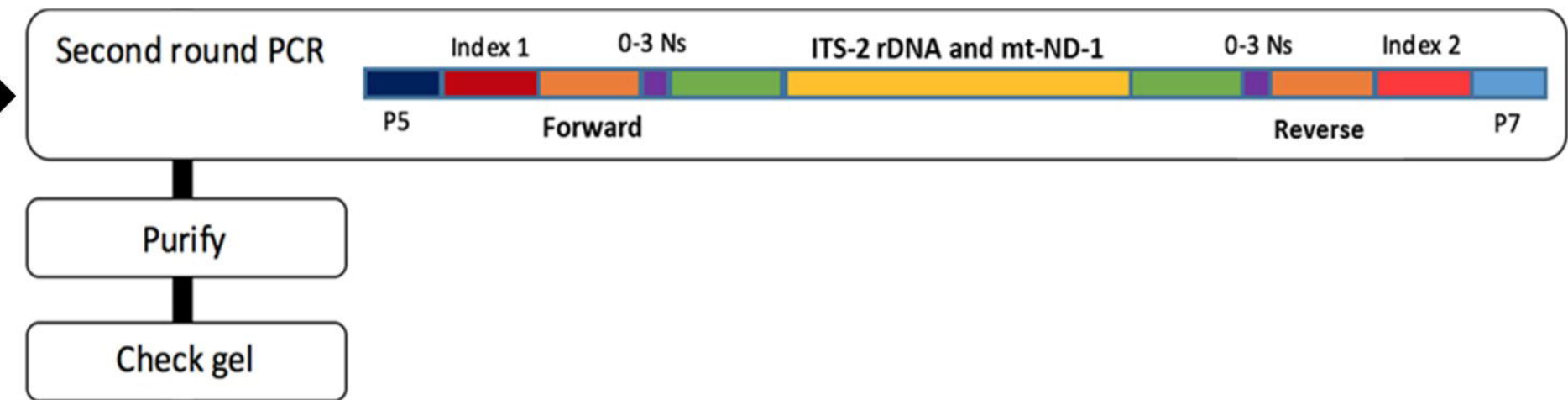
Supplementary Figure Legends

Supplementary Fig. S1. Maximum likelihood tree of twenty-six mt-ND-1 haplotypes sequenced from twenty *F. gigantea* populations from two provinces of Pakistan as indicated on the insert table. The tree was constructed in Mega 6.0 (Tamura et al., 2013) based on a neighbor-joining algorithm using HKY+G+I model. The tree was rooted with the corresponding mt-ND-1 sequence of *F. hepatica* (4P01770.1), and the branch supports were obtained by 1000 bootstraps of the data. The colored rectangles and circles indicate the populations in which each haplotype was identified as indicated by the color key in the insert table.

Amplification of rDNA ITS-2 and mt-ND-1 locus using adapter primers



Amplicons generated from each sample with barcoded PCR



Bioinformatics data handling

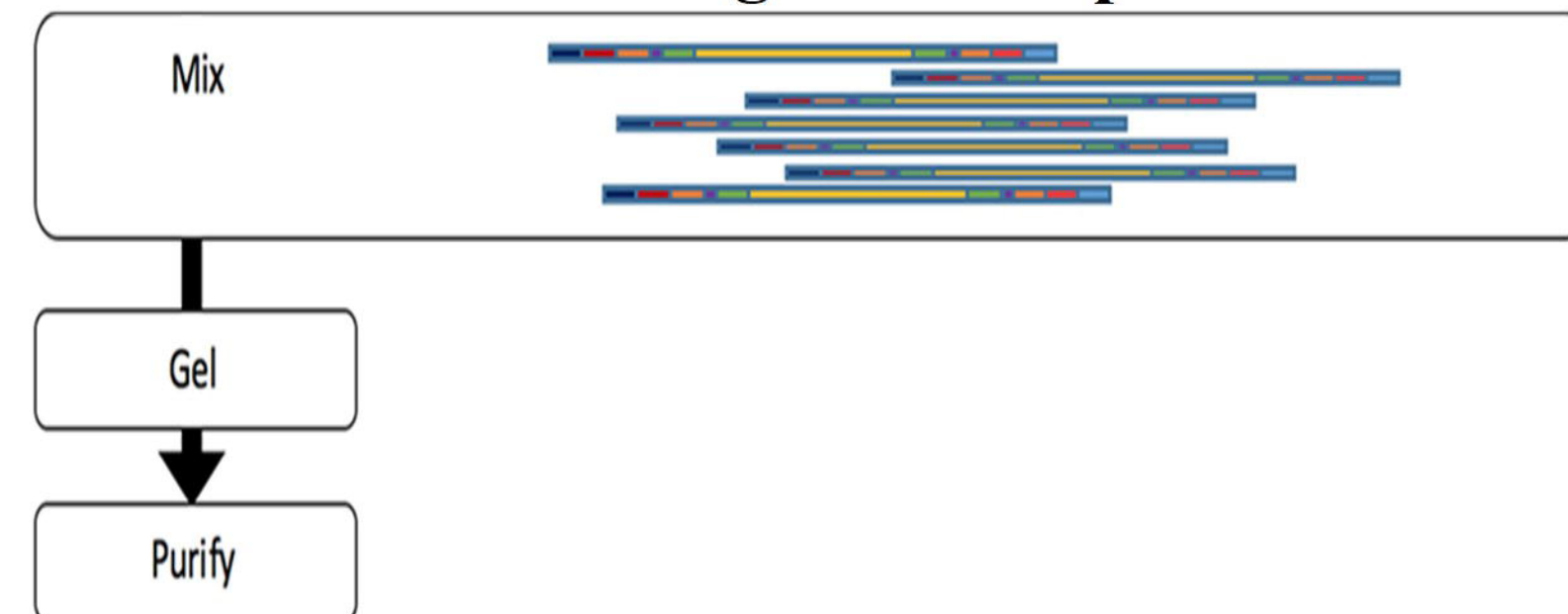
```
File Edit View Scrollback Settings Help
[root@txxmachine]# ls Pictures/ -Salh
total 1.2M
-rw-rw-r-- 1 s s 190K 2009-11-06 10:16 sl_dlink_support.png
-rw-rw-r-- 1 s s 137K 2009-11-06 10:50 sl_adv10_kdeccrpadx.png
-rw-rw-r-- 1 s s 124K 2009-11-06 10:32 sl_dlink_wrisscrtx.png
-rw-rw-r-- 1 s s 119K 2009-11-05 22:00 sl_adv10kdeccrpadrx.png
-rw-rw-r-- 1 s s 116K 2009-11-06 00:23 sl_adv10_gnomeinstaller.png
-rw-rw-r-- 1 s s 102K 2009-11-06 10:16 sl_dlink_support.gif
-rw-rw-r-- 1 s s 72K 2009-11-06 10:41 sl_dlink_security2.png
-rw-rw-r-- 1 s s 67K 2009-11-06 10:20 sl_dlink_support.jpg
-rw-rw-r-- 1 s s 63K 2009-11-04 21:26 codeina.png
-rw-rw-r-- 1 s s 54K 2009-11-06 10:53 sl_adv10_kde.png
-rw-rw-r-- 1 s s 15K 2009-11-06 10:53 sl_dlink_wrisscrtx.png
shad-xr-x 2 s s 4.0K 2009-11-06 11:03 .../
shad-xr-x 39 s s 4.0K 2009-11-06 11:03 .../
-rw-rw-r-- 1 s s 15 2009-11-06 00:53 sl_dlink_wrisscrtx.png
[root@txxmachine]#
[root@txxmachine]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda7        20G   8.5G   11G   46% /
/dev/sdb15       9.7G   8.2G   1.1G   89% /mnt/sdb15
[root@txxmachine]#
```

Command prompt
command line /
mother

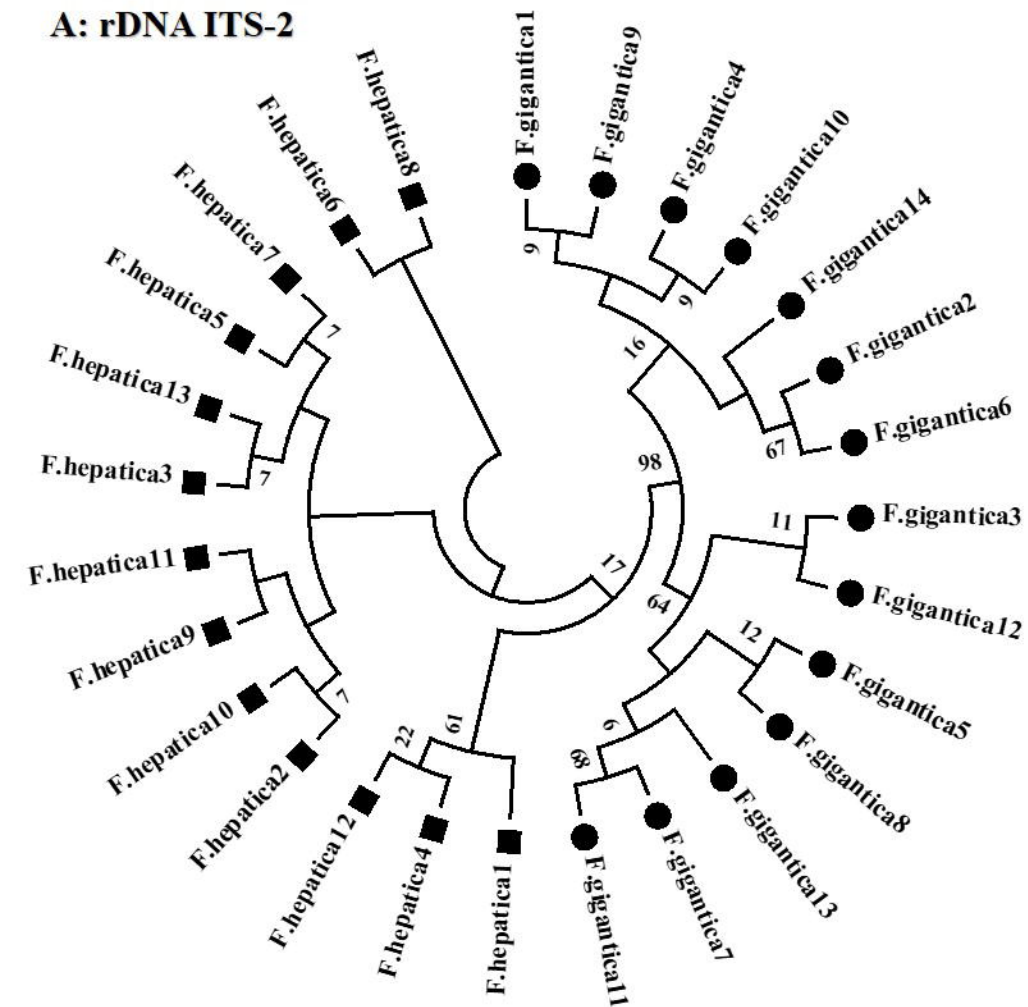
Sequencing each sample on Illumina MiSeq platform



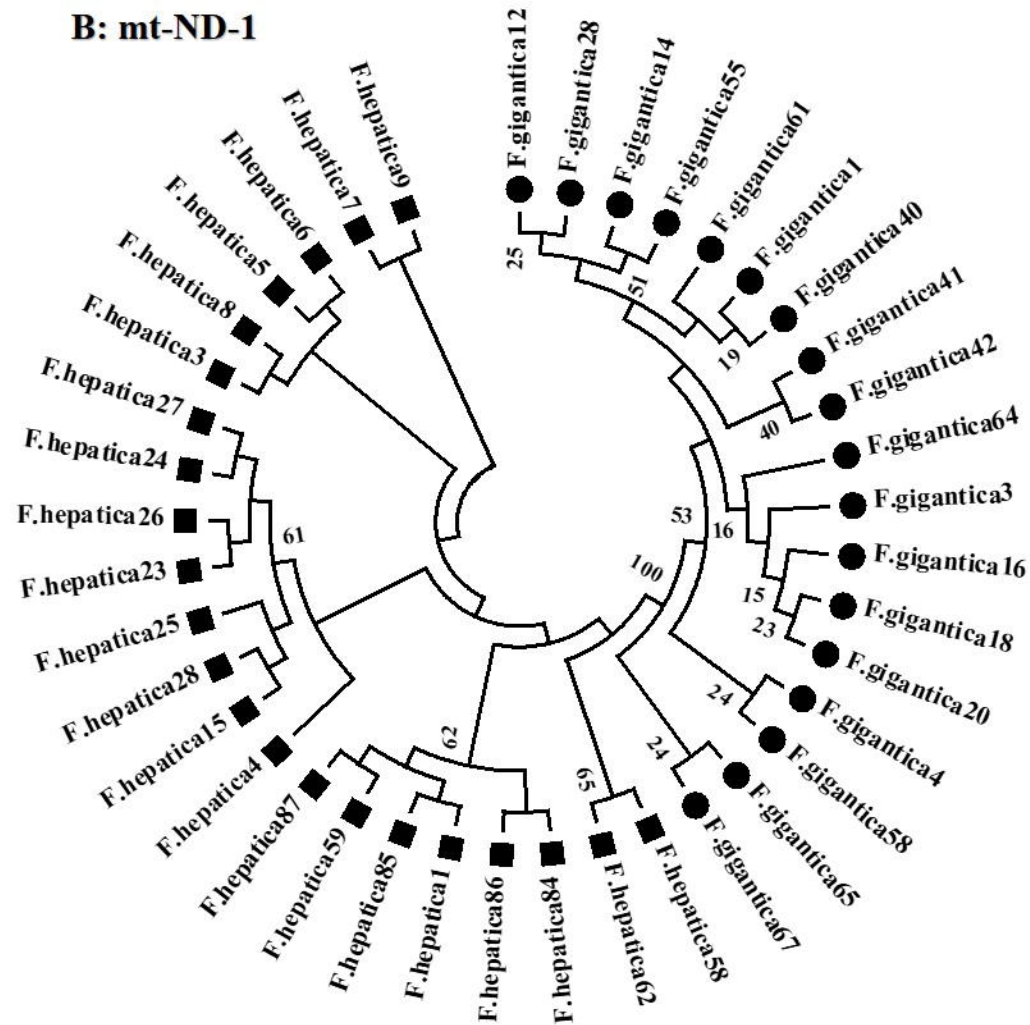
Pooling of all samples

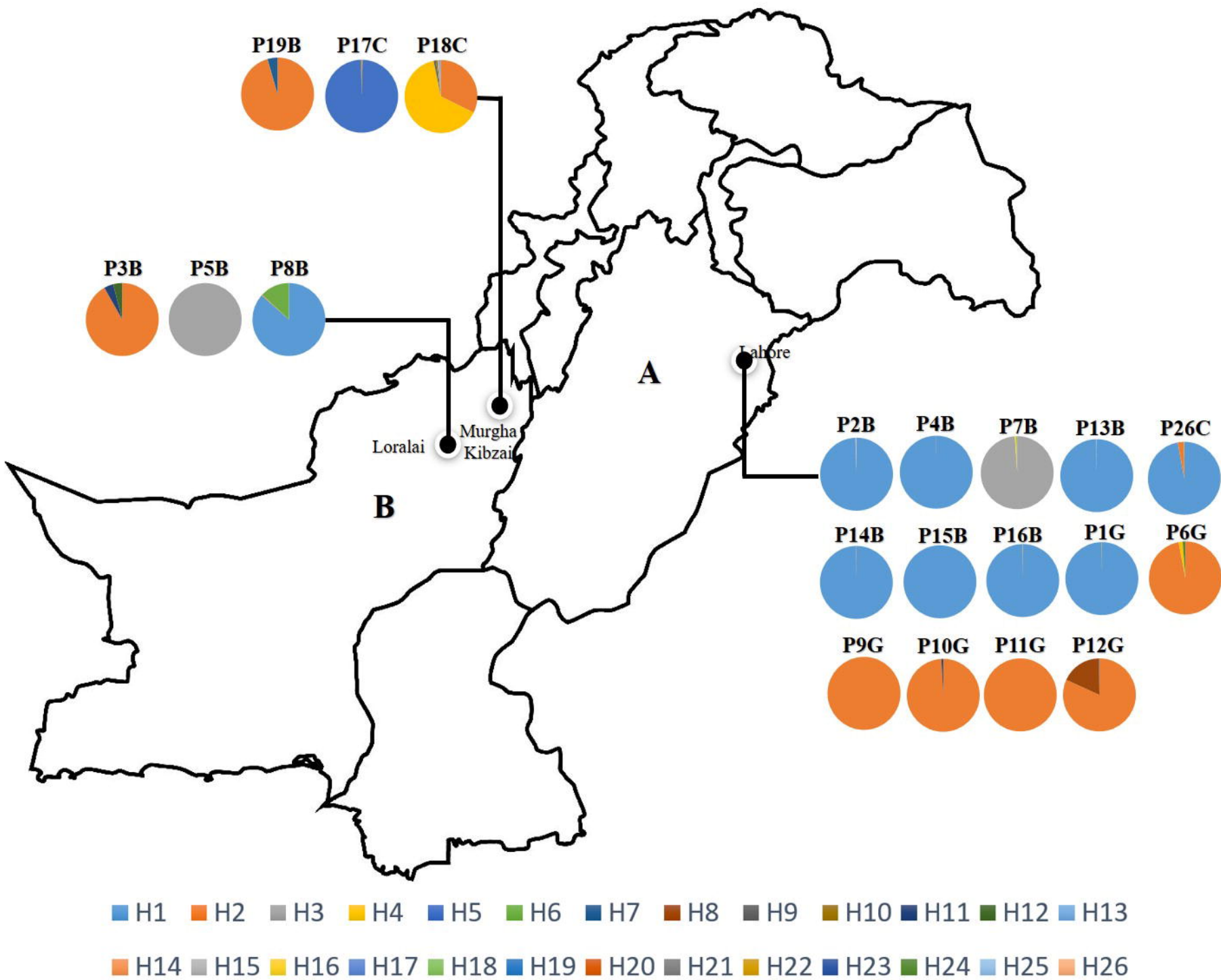


A: rDNA ITS-2

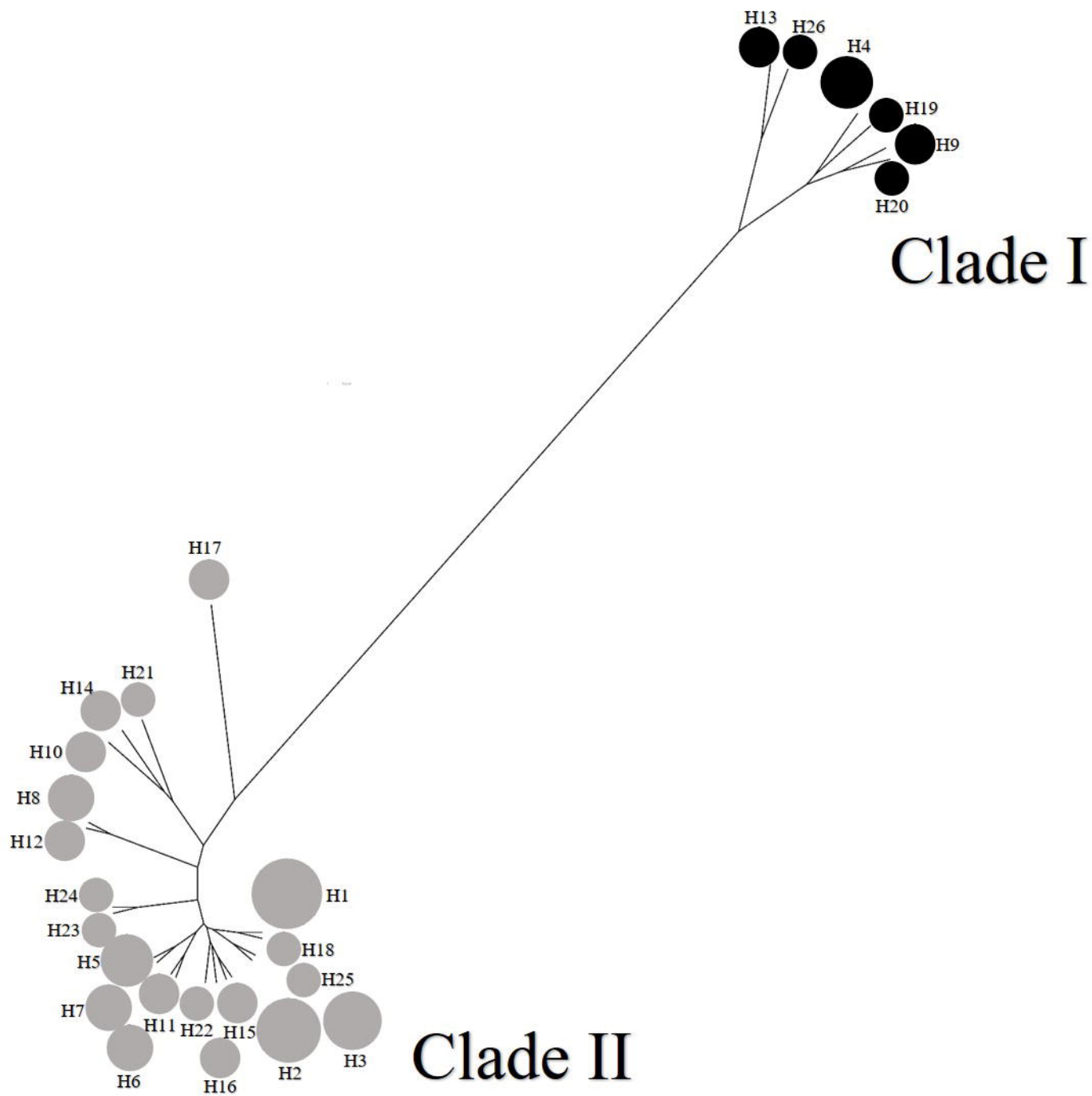


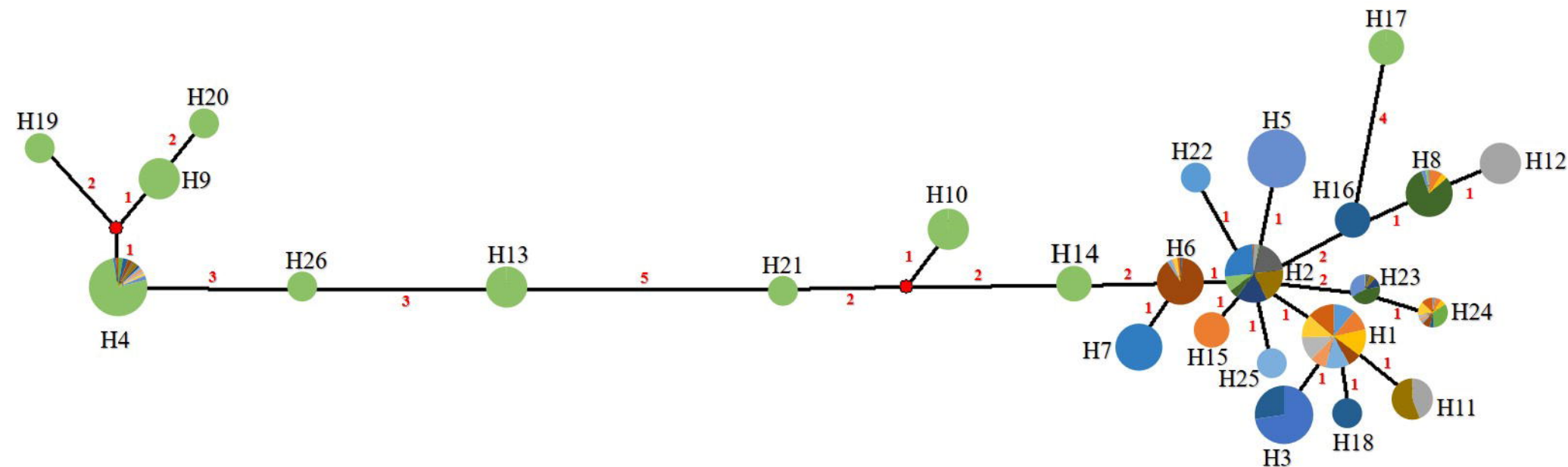
B: mt-ND-1





Populations	Sequence reads (%)	Predominated haplotypes (%)	Rare haplotypes (%)
P1G	82376(5.531)	H1(99.83005)	H22(0.166311), H24(0.003642)
P2B	82495(5.539)	H1(99.51027)	H2(0.098188), H8(0.018183), H14(0.363658), H15(0.343658), H16(0.264672), H24(0.009698)
P3B	12776(0.857)	H2(91.8206)	H11(4.15623), H12(4.64325), H13(4.023168)
P4B	10532(7.072)	H1(99.94303)	H2(0.021838), H8(0.028485), H24(0.006646)
P5B	12240(8.219)	H5(100)	
P6G	13237(0.217)	H2(97.0034)	H4(1.729997), H10(0.061786), H24(1.204819),
P7B	46363(3.113)	H7(98.92587)	H4(0.023726), H6(0.012941), H17(0.556478), H18(0.470203), H24(0.010784)
P8B	58506(3.928)	H1(86.53813), H6(13.12515)	H2(0.295696), H4(0.020511), H24(0.017092)
P9G	79854(5.362)	H2(99.98497)	H4(0.005009), H23(0.010018)
P10G	81470(5.471)	H2(99.15797)	H4(0.00491), H11(0.827298), H12(0.746678), H23(0.00982)
P11G	71509(4.802)	H2(99.97623)	H4(0.004195), H23(0.019578)
P12G	24854(1.669)	H2(81.78563), H8(17.97296)	H23(0.24141)
P13B	92620(6.219)	H1(99.84129)	H2(0.029151), H4(0.006478), H6(0.004319), H24(0.003239), H25(0.115526)
P14B	62230(4.179)	H1(99.8843)	H2(0.096417), H4(0.011249), H6(0.001607), H24(0.006428)
P15B	92220(6.192)	H1(99.96855)	H2(0.01735), H4(0.007591), H8(0.001084), H24(0.005422)
P16B	89654(6.020)	H1(99.85165)	H2(0.107078), H4(0.016731), H6(0.005577), H24(0.018962)
P17C	57043(3.830)	H5(99.28826)	H2(0.082394), H3(0.026296), H4(0.238417), H8(0.287502), H23(0.077135)
P18C	10428(7.002)	H4(64.47613), H2(32.25389)	H8(0.002877), H9(0.998255), H10(0.703861), H19(0.196582), H20(0.165896), H21(0.154389), H23(0.000959)
P19B	11258(7.560)	H2(95.47732)	H3(0.013323), H4(0.046186), H7(4.435641), H8(0.027534)
P26C	10730(7.206)	H1(97.08503)	H2(2.793827), H4(0.030753), H6(0.076416), H24(0.013978)





Lahore (Punjab)				
P1G	P6G	P10G	P13B	P16B
P2B	P7B	P11G	P14B	P26C
P4B	P9G	P12G	P15B	

Loralai (Balochistan)
P3B
P5B
P8B

Murgha Kibzai (Balochistan)
P17C
P18C
P19B

Haplotypes	Sequence reads (%)	No. of populations
H1	652740(43.825)	9 (P1G, P2B, P4B, P8B, P13B, P14B, P15B, P16B, P26C)
H2	425968(28.599)	17 (P2B, P3B, P4B, P6G, P8B, P9G, P10G, P11G, P12G, P13B, P14B, P15B, P16B, P17C, P18C, P19B, P26C)
H3	198298(13.313)	4 (P5B, P7B, P17C, P19B)
H4	69583(4.671)	14 (P6G, P7B, P8B, P9G, P10G, P11G, P13B, P14B, P15B, P16B, P17C, P18C, P19B, P26C)
H5	89637(6.018)	1 (P17C)
H6	4777(0.320)	6 (P7B, P8B, P13B, P14B, P16B, P26C)
H7	4194(0.281)	1 (P19B)
H8	4011(0.269)	7 (P2B, P4B, P12G, P15B, P17C, P18B, P19B)
H9	3741(0.251)	1 (P18C)
H10	3636(0.244)	2 (P6G, P18C)
H11	3805(0.255)	2 (P3B, P10G)
H12	3414(0.229)	1 (P3B)
H13	3417(0.229)	2 (P8B, P18C)
H14	2930(0.196)	1 (P18C)
H15	3010(0.202)	1 (P2B)
H16	2958(0.198)	1 (P7B)
H17	2844(0.190)	2 (P8B, P18C)
H18	1218(0.081)	1 (P7B)
H19	1305(0.087)	1 (P18C)
H20	1173(0.078)	1 (P18C)
H21	1161(0.077)	1 (P18C)
H22	1137(0.076)	1 (P1G)
H23	1135(0.076)	6 (P9G, P10G, P11G, P12G, P17C, P18C)
H24	1116(0.074)	11 (P1G, P2B, P4B, P6G, P7B, P8B, P13B, P14B, P15B, P16B, P26C)
H25	1107(0.074)	1 (P13B)
H26	1103(0.074)	1 (P18C)

Table 1. Species identification of individual worms. from twenty-six *Fasciola* populations were performed. Each population represents the flukes collected from a single host. The table is divided into two sections based on the province from where the samples were collected.

Populations	Host	Area	Province	Total flukes	<i>F. gigantica</i>	<i>F. hepatica</i>	Heterozygous
P2B	Buffalo	Lahore	Punjab	16	16	-	-
P4B	Buffalo	Lahore	Punjab	15	15	-	-
P7B	Buffalo	Lahore	Punjab	8	8	-	-
P13B	Buffalo	Lahore	Punjab	20	20	-	-
P14B	Buffalo	Lahore	Punjab	20	20	-	-
P15B	Buffalo	Lahore	Punjab	20	20	-	-
P16B	Buffalo	Lahore	Punjab	20	20	-	-
P26C	Cattle	Lahore	Punjab	20	20	-	-
P1G	Goat	Lahore	Punjab	4	4	-	-
P6G	Goat	Lahore	Punjab	16	16	-	-
P9G	Goat	Lahore	Punjab	7	7	-	-
P10G	Goat	Lahore	Punjab	6	6	-	-
P11G	Goat	Lahore	Punjab	4	4	-	-
P12G	Goat	Lahore	Punjab	7	7	-	-
P3B	Buffalo	Loralai	Balochistan	11	11	-	-
P5B	Buffalo	Loralai	Balochistan	12	12	-	-
P8B	Buffalo	Loralai	Balochistan	7	7	-	-
P17C	Cattle	Murgha Kipzai	Balochistan	12	3	6	3
P18C	Cattle	Murgha Kipzai	Balochistan	12	8	2	2
P19C	Cattle	Murgha Kipzai	Balochistan	18	8	7	3
P20C	Cattle	Murgha Kipzai	Balochistan	20	3	16	1
P21S	Sheep	Murgha Kipzai	Balochistan	15	3	11	1
P22S	Sheep	Murgha Kipzai	Balochistan	3	-	3	-
P23S	Sheep	Loralai	Balochistan	1	-	1	-
P24S	Sheep	Loralai	Balochistan	3	-	3	-
P25S	Sheep	Loralai	Balochistan	8	-	7	1
Total				305	238	56	11

Table 2: Genetic diversity estimation of mt-ND-1 haplotypes identified from twenty populations of *Fasciola gigantica* in Pakistan. The table shows these values both with populations and the total between populations.

Populations	No. of Illumina MiSeq reads	No. of haplotype generated	Haplotype diversity (H_d)	Segregating sites (S)	Nucleotide diversity (Π)	Mutation parameter based on S (Θ_s)	Mean No. of pairwise difference (k)
P1G	82376	3	0.750	4	0.00643	0.00473	2.000
P2B	82495	5	0.857	7	0.00827	0.00692	2.571
P3B	12776	3	0.727	6	0.00935	0.00639	2.909
P4B	10532	4	0.800	6	0.00815	0.00581	2.533
P5B	12240	1			N/A		
P6G	13237	4	0.818	21		0.02236	9.136
P7B	46363	6	0.882	22	0.02326	0.02057	7.235
P8B	58506	7	0.900	21	0.02811	0.01877	8.743
P9G	79854	3	0.727	20	0.03118	0.02130	9.697
P10G	81470	4	0.818	21	0.02806	0.02236	8.727
P11G	71509	3	0.727	20	0.03118	0.02130	9.697
P12G	24854	3	0.727	5	0.00779	0.00532	2.424
P13B	92620	6	0.882	22	0.02194	0.02057	6.824
P14B	62230	5	0.857	21	0.02425	0.02077	7.543
P15B	92220	5	0.857	22	0.02591	0.02176	8.057
P16B	89654	5	0.857	21	0.02425	0.02077	7.543
P17C	57043	6	0.882	24	0.02383	0.02244	7.412
P18C	10428	13	0.960	21	0.03055	0.01770	9.502
P19B	11258	5	0.857	22	0.02591	0.02176	8.057
P26C	10730	5	0.857	21	0.02425	0.02077	7.543
Total	1002395	26	0.917	24	0.02466	0.01628	7.670

